Prediction of in-hospital Mortality of Intensive Care Unit Patients with Acute Pancreatitis Based on an Explainable Machine Learning Algorithm

Wensen Ren, MD,† Kang Zou, MD,*† Shu Huang, MD, PhD,‡§ Huan Xu, MD,*† Wei Zhang, MD, PhD,*† Xiaomin Shi, MD, PhD,*† Lei Shi, MD, PhD,*† Xiaolin Zhong, MD, PhD,*† Yan Peng, MD, PhD,*† Xiaowei Tang, MD, PhD,*† and Muhan Lü, MD, PhD*†*

Background and Aim: Acute pancreatitis (AP) is potentially fatal. Therefore, early identification of patients at a high mortality risk and timely intervention are essential. This study aimed to establish an explainable machine-learning model for predicting in-hospital mortality of intensive care unit (ICU) patients with AP.

Methods: Data on patients with AP, including demographics, vital signs, laboratory tests, comorbidities, treatment, complication, and severity scores, were extracted from the Medical Information Mart for Intensive Care IV (MIMIC-IV) and the eICU collaborative research database (eICU-CRD). Based on the data from MIMIC-IV, we used the least absolute shrinkage and selection operator algorithm to select variables and then established 9 machine-learning models and screened the optimal model. Data from the eICU-CRD were used for external validation. The area under the receiver operating characteristic curve (AUC), sensitivity, specificity, accuracy, decision curve, and calibration plots were used to assess the models' efficacy. Shapley's additive explanation values were used to explain the model.

Results: Gaussian naive Bayes (GNB) model had the best performance on the data from MIMIC-IV, with an AUC, accuracy, sensitivity, and specificity of 0.840, 0.787, 0.839, and 0.792, respectively. The GNB model also performed well on the data from the eICU-CRD, with an AUC, accuracy, sensitivity, and specificity of 0.862, 0.833, 0.848, and 0.763, respectively. According to Shapley's additive explanation values, the top 4 predictive factors were maximum red cell distribution width, minimum saturation of blood oxygen, maximum blood urea nitrogen, and the Sequential Organ Failure Assessment score.

Conclusion: The GNB model demonstrated excellent performance and generalizability in predicting mortality in ICU patients with AP. Therefore, it can identify patients at a high mortality risk.

Key Words: acute pancreatitis, intensive care unit, machine learning

(*J Clin Gastroenterol* 2024;58:619–626)

Acute pancreatitis (AP) is a self-digestive disease of the pancreas,[1] and its incidence is increasing worldwide.[2] In the United States, AP is the main cause of hospitalization for gastrointestinal diseases and incurs an annual cost of ~ $2.5 billion.[3,4] According to the Revised Atlanta Classification, organ failure and disease duration are important indicators of disease severity. However, there is no organ failure in mild AP cases, transient organ failure in moderately severe AP cases, or persistent organ failure in severe AP cases.[5] Most patients with AP exhibit mild symptoms, and the disease course is self-limiting. Approximately 20% of patients experience organ failure, and their condition develops into moderate or severe AP, with a mortality rate of ~20% to 40%.[6] Studies have shown that the sequence and number of organ failures affect mortality in patients with AP. Compared with other systems, early involvement of the cardiovascular and respiratory systems results in worse outcomes.[7] Biochemical markers (interleukin-6, C reactive protein, procalcitonin, serum albumin, and growth differentiation factor 15) and Scoring systems [Ranson's score, Glasgow score, multiple organ system score, systemic inflammatory response syndrome, bedside index for severe acute pancreatitis, acute physiology and chronic health evaluation II (APACHE-II), computed tomography severity index, and sequential organ failure assessment (SOFA) score] were used to predict the severity and prognosis of AP.[8–11]

With the development of artificial intelligence, machine learning is increasingly used in the field of clinical medicine. A previous study reported that an artificial neural network (ANN) model demonstrated excellent performance in predicting fatal outcomes with an area under the receiver operating characteristic curve (AUC) value of 0.847.[12] However, the study did not explain the choices made by the model. Therefore, the present study aimed to establish an explainable machine-learning model for predicting in-hospital mortality of ICU patients with AP.

Received for publication April 26, 2023; accepted July 16, 2023.
From the *Department of Gastroenterology, the Affiliated Hospital of Southwest Medical University; †Nuclear Medicine and Molecular Imaging Key Laboratory of Sichuan Province, Luzhou; ‡Department of Gastroenterology, Lianshui County People's Hospital; and §Department of Gastroenterology, Lianshui People's Hospital of Kangda College Affiliated to Nanjing Medical University, Huaian, China.

W.R., K.Z., and S.H. contributed equally to this study.

W.R., X.T., S.H., and K.Z: Study conception and design; W.R., X.T. and K.Z.: Drafting of the manuscript; W.Z., X.S., L.S., X.Z., and Y.P.: Acquisition of data and critical revision; M.L: Revision of manuscript, and final approval of manuscript.

This work is supported by the Natural Science Foundation of Sichuan Province (No 2022NSFSC1378), and the Extreme Smart Analysis platform (http://xsmartanalysiscn/).

The authors declare that they have nothing to disclose.

Address correspondence to: Muhan Lü, MD, PhD, Department of Gastroenterology, the Affiliated Hospital of Southwest Medical University, Street Taiping No.25, Region Jiangyang, Luzhou, 646099, Sichuan Province, China (e-mail: lvmuhan@swmu.edu.cn).

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.jcge.com.

DOI: 10.1097/MCG.0000000000001910

METHODS

Data Source

We extracted data from the Medical Information Market for Intensive Care (MIMIC-IV) and eICU collaborative research databases (eICU-CRD). MIMIC-IV, a single-center database, included clinical data from > 190000 patients and > 450000 hospitalization records of patients admitted to the Beth Israel Deaconess Medical Center between 2008 and 2019. The eICU-CRD included information on 335 ICU patients from 208 hospitals across the United States, including > 200000 patients admitted to the ICU between 2014 and 2015. Both databases included large amounts of high-quality clinical data. The ethics committee of the affiliated hospital of Southwest Medical University abandoned the pursuit of ethics for the privacy protection policies of the databases.

Participants

Patients diagnosed with AP admitted to the ICU were enrolled in this study. Exclusion criteria were:(1) not the first ICU record, (2) ICU stay of <24 hours, and (3) age <18 years. Finally, 1281 patients were enrolled in this study, of whom 856 were from MIMIC-IV and 425 were from the eICU-CRD (Fig. 1).

Data Collection and Outcome Measurement

We retrieved patient data in the initial 24 hours after admission to the ICU from the MIMIC-IV and eICU-CRD. Information on demographic data (sex, age, weight, and height), vital signs [heart rate, respiratory rate(RR), temperature(TP), oxygen saturation(SpO2), systolic blood pressure(SBP), diastolic blood pressure(DBP), and mean arterial pressure(MBP)], comorbidities (hypertension and diabetes), complication [acute kidney injury(AKI)], prognostic scoring system [Oxford acute severity of illness score (OASIS) and SOFA], treatment [Mechanical ventilation and Renal replacement therapy (RRT)], laboratory tests [Sodium, Chloride, Bicarbonate, Potassium, Calcium, Anion gap, Creatinine, blood urea nitrogen (BUN), Albumin, Bilirubin, aminotransferase aspartate (AST), aminotransferase alanine (ALT), platelets, white blood cells (WBC), red cell distribution width (RDW), red blood cells (RBC), Hemoglobin, and Hematocrit] were collected. Survival status at discharge was set as the study outcome. Considering that variables may deviate from the normal range, indicating a pathologic state, we extracted the maximum, minimum, and mean values for analysis. Because bias increases with an increase in missing values, we excluded variables with missing values exceeding 30%.

Predictor Selection

The least absolute shrinkage and selection operator was used to select variables based on variables with differences between survival and nonsurvival group in MIMIC-IV (Fig. 2). The least absolute shrinkage and selection operator algorithm incorporates a penalty function to minimize the occurrence of collinearity and overfitting. Finally, 11 variables were included for model construction (Age, AKI, SBP_mean, SPO2_min, RDW_max, WBC_max, BUN_max, Albumin_min, Bilirubin_total_max, OASIS, SOFA)

Model Development

In the model development phase, we constructed 9 predictive models, namely XGBoost, LightGBM, Random Forest, AdaBoost, Gaussian naive Bayes (GNB), complement naive Bayes, multi-layer perceptron, support vector machine, and K-nearest neighbors, using data from MIMIC-IV. To fully use the data information, we used the 10-fold cross-verification method to obtain the average performance of the model. AUC, sensitivity, specificity, and accuracy were used to evaluate performance. To evaluate the generalizability of the best model for MIMIC-IV, we used data from the eICU-CRD for external verification. The features' importance (Supplementary Figure 1, Supplemental Digital Content 1, http://links.lww.com/JCG/A1000) and Shapley's additive explanation (SHAP) plots (Fig. 5) were used to explain the decisions made by the model.

Statistical Analysis

Continuous variables were expressed as mean ± SD or median with interquartile range and were tested using

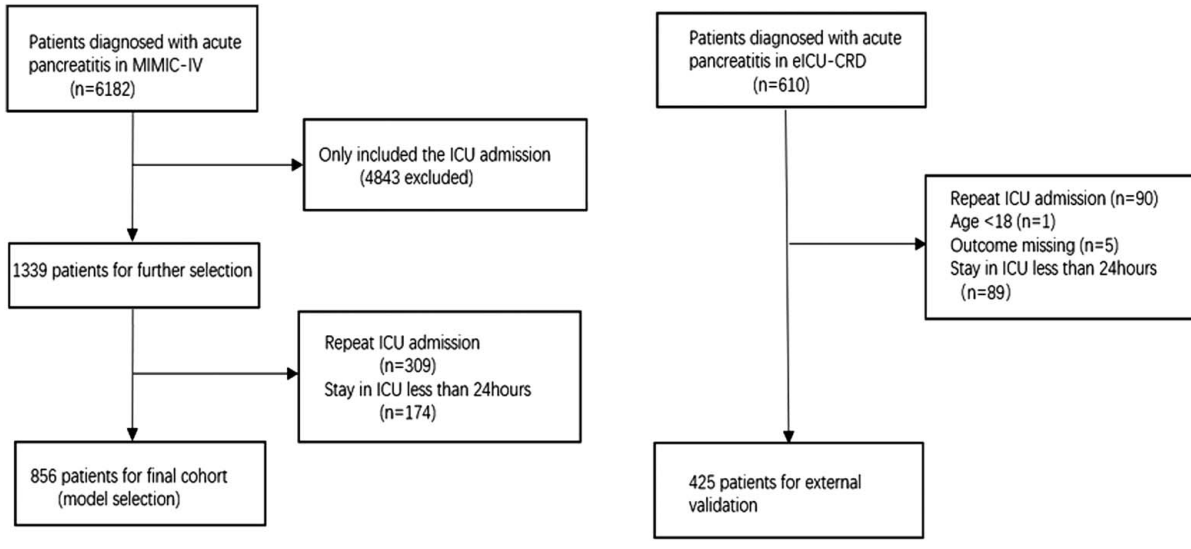

FIGURE 1. Flowchart of patients' selection.

Copyright © 2023 Wolters Kluwer Health, Inc. All rights reserved.

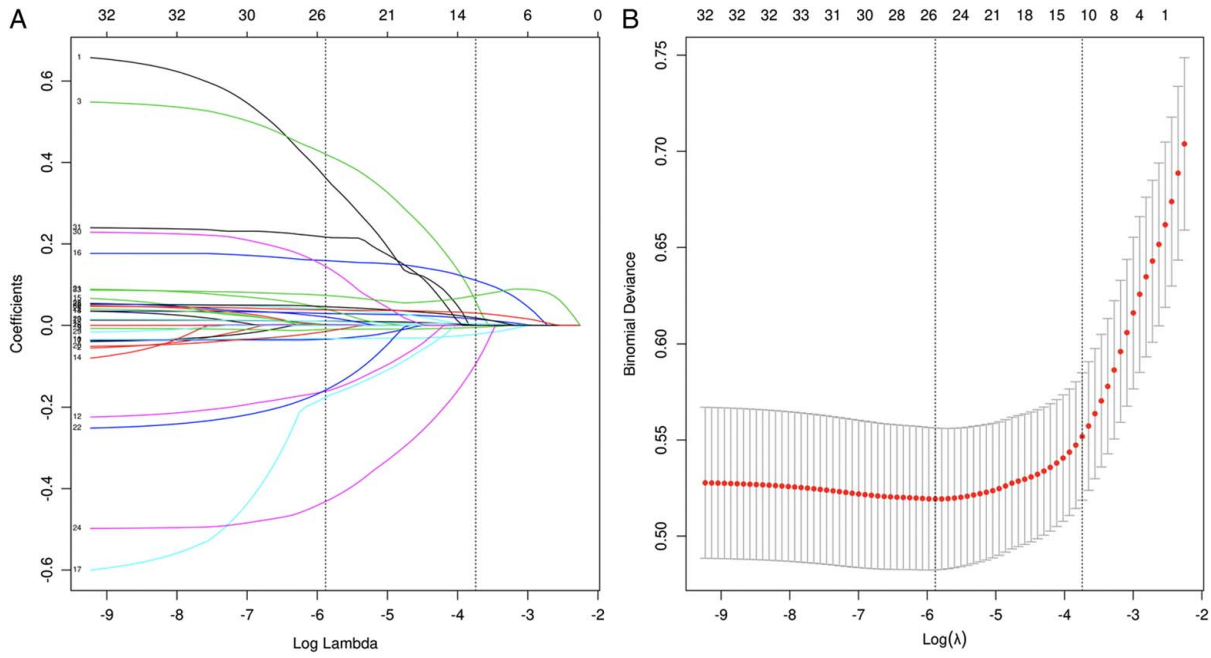

FIGURE 2. Clinical feature selection based on the LASSO logistic regression. (A) Selection of the Optimal lambda in the LASSO logistic regression. Each line represents the change of coefficient of each feature. (B) LASSO coefficient profiles of features. The left and right black vertical lines were drawn at the lambda with minimum deviance and 1 SE to the lambda with minimum deviance. LASSO indicates least absolute shrinkage and selection operator.

Student *t*-test or Mann–Whitney *U* test. Categorical variables were expressed as counts and percentages and were tested using the χ^2 test. We used the multivariate imputation method for <30% of the missing variables. All Statistical analyses were performed using R version 3.6.3 and Python version 3.7

RESULTS

Patient Characteristics

We identified 856 and 425 patients from the MIMIC-IV and eICU-CRD, respectively. The in-hospital mortality rates in the MIMIC-IV and eICU-CRD were 12.97% and

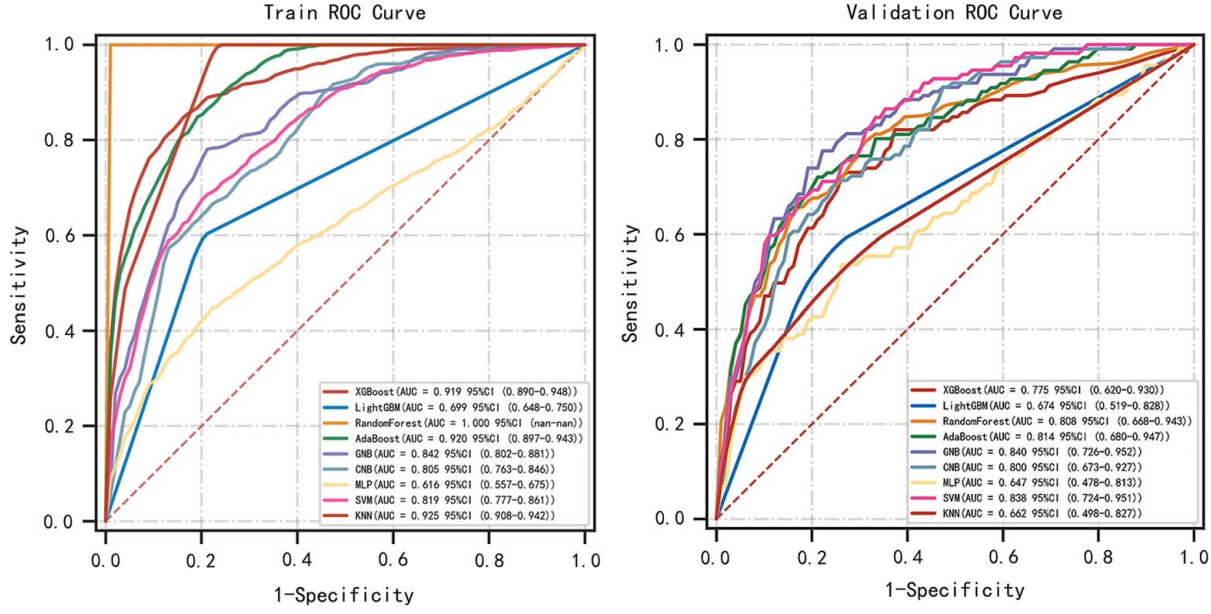

FIGURE 3. ROC curve of models in training and validation set. GNB indicates Gaussian naive bayes; CNB, complement naive bayes; KNN, K-nearest neighbors; MLP, multi-layer perceptron; ROC, receiver operating characteristic; SVM, support vector machine.

TABLE 1. Prediction Performance of the Machine Learning Models in the Validation Set

Classifiers	AUC (SD)	Accuracy (SD)	Sensitivity (SD)	Specificity (SD)
XGBoost	0.775 (0.620–0.930)	0.776 (0.752–0.799)	0.758 (0.663–0.852)	0.775 (0.716–0.834)
LightGBM	0.674 (0.519–0.828)	0.870 (0.868–0.873)	0.561 (0.474–0.647)	0.787 (0.762–0.811)
Random Forest	0.808 (0.668–0.943)	0.881 (0.865–0.896)	0.828 (0.755–0.901)	0.734 (0.617–0.850)
AdaBoost	0.814 (0.680–0.947)	0.765 (0.734–0.796)	0.757 (0.657–0.856)	0.817 (0.750–0.884)
GNB	0.840 (0.726–0.952)	0.787 (0.765–0.810)	0.839 (0.779–0.899)	0.792 (0.726–0.858)
CNB	0.800 (0.673–0.927)	0.688 (0.622–0.754)	0.792 (0.721–0.863)	0.762 (0.670–0.854)
MLP	0.647 (0.478–0.813)	0.733 (0.675–0.792)	0.658 (0.510–0.805)	0.713 (0.533–0.893)
SVM	0.838 (0.724–0.951)	0.762 (0.717–0.806)	0.873 (0.819–0.928)	0.726 (0.643–0.808)
KNN	0.662 (0.498–0.827)	0.854 (0.832–0.876)	0.480 (0.353–0.606)	0.832 (0.755–0.910)

CNB indicates Complement naive bayes; GNB, Gaussian naive bayes; KNN, K-nearest neighbors; MLP, Multi-layer perceptron; SVM, support vector machine.

7.76%, respectively. Age, SPO2_min, SPO2_mean, TP_max, TP_min, MBP_mean, DBP_mean, SBP_mean, RR_mean, RR_max, Bicarbonate_max, Bicarbonate_min, Potassium_max, Potassium_min, Calcium_min, Aniongap_max, Aniongap_min, Creatinine_max, BUN_max, Albumin_min, Bilirubin_total_max, AST, Platelets_min, WBC_max, RDW_max, RBC_min, Hemoglobin_min, Hematocrit_min, AKI, OASIS, SOFA, Mechanical ventilation, and RRT differed between the nonsurvivors and survivors group in MIMIC-IV, whereas Age, MBp_mean, DBP_mean, SBP_mean MBP_mean, DBP_mean, SBP_mean, RR_mean, RR_max, Creatinine_max, BUN_max, Albumin_min, Bilirubin_total_max, WBC_max, Rdw_max, Rbc_min, Hemoglobin_min, Hematocrit_min, Hypertension, AKI, OASIS, SOFA, Mechanical ventilation, and RRT differed between the nonsurvivors and survivors group in the eICU-CRD.

Development Models For MIMIC-IV

Regarding MIMIC-IV, we developed 9 machine-learning prediction models and drew an AUC curve (Fig. 3) to display the performance of the models. The random Forest model performed excellently on the training set; however, the GNB model performed better on the validation set. As shown in Table 1, we considered the GNB model with AUC, accuracy, sensitivity, and specificity of 0.840, 0.787, 0.839, and 0.792, respectively, the best model because the Random Forest model has a risk of overfitting.

External Validation of the GNB Model

To evaluate the generalizability of the GNB model, we drew an AUC curve (Fig. 4), calibration plots (Supplementary Figure 2, Supplemental Digital Content 1, http://links.lww.com/JCG/A1000), and a decision curve (Supplementary Figure 3, Supplemental Digital Content 1, http://links.lww.com/JCG/A1000) using the eICU-CRD. Finally, the model demonstrated comparable performance on the eICU-CRD database, with an AUC value, accuracy, sensitivity, and specificity of 0.862, 0.833, 0.848, and 0.763, respectively. RDW_max, SPO2_min, BUN_max, and SOFA scores were the 4 most important factors affecting the model's decision (Supplementary Figure 1, Supplemental Digital Content 1, http://links.lww.com/JCG/A1000). According to the SHAP plot (Fig. 5), lower SPO2_min and higher RDW_max, BUN_max, and SOFA scores indicated a worse prognosis.

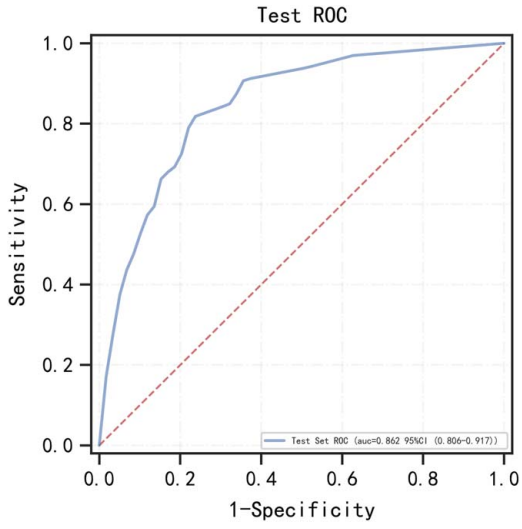

FIGURE 4. ROC curve of Gaussian NB model in the test set. ROC indicates receiver operating characteristic.

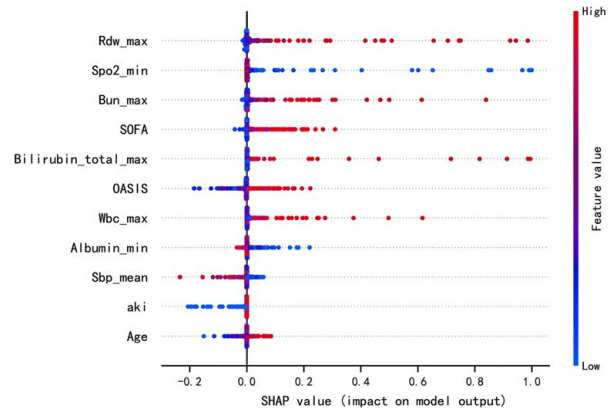

FIGURE 5. SHAP plot of features of GNB model in the test set. The colors reflect the value of variables. The SHAP value represents the correlation with death, with a positive value indicating a risk of death and a negative value indicating the opposite. aki indicates acute kidney injury; Bun, blood urea nitrogen; OASIS, Oxford acute severity of illness score; Rdw, red cell distribution width; ROC, operating characteristic; Sbp, systolic blood pressure; Spo2, oxygen saturation; SOFA, sequential organ failure assessment; Wbc, white blood cells.

Copyright © 2023 Wolters Kluwer Health, Inc. All rights reserved.

DISCUSSION

AP is a common digestive system disease with severity varying from mild to severe with different prognoses.[5] Organ failure is a significant factor that affects mortality.[6] Patients with mild AP without organ failure mostly recover in the first week, and their mortality rate is low.[5] However, the mortality rate of patients with severe AP with persistent organ failure is as high as 20% to 40%.[6] Therefore, reducing the mortality rate of patients with high-risk AP was the focus of our efforts.

Patients admitted to the ICU are in critical condition. In this study, we established several machine-learning models to predict the clinical outcomes of patients with AP admitted to the ICU. The universal applicability of a model is an important index for evaluating its quality. The Random Forest model performed excellently on the training set; however, its performance declined rapidly on the validation set. This model is believed to have the risk of overfitting, which may affect its universal applicability. Therefore, we believe the GNB model is better for achieving stable performance. To further verify the generalizability of the GNB model, the eICU-CRD was used for external validation. As shown in Fig. 4, the GNB model demonstrated comparable performances, with an AUC value of 0.863, on the external validation database, indicating that the GNB model has universal applicability.

According to the SHAP values, we sorted the variables' importance and explained the model's decision (Supplementary Figure 1, Supplemental Digital Content 1, http://links.lww.com/JCG/A1000 and Fig. 5). Spo2_min was considered the important predictor of the GNB model. Moderate-to-severe patients experience organ failure.[5] A previous study reported respiratory failure as the most common organ failure.[13] In their study, 221 out of 240 patients with AP experienced respiratory failure, with a mortality rate of 37%.[13] In the present study, 39.14% and 28.00% of the patients underwent mechanical ventilation in the MIMIC-IV database and eICU-CRD, respectively. Notably, the proportion of patients who underwent mechanical ventilation was much higher in the mortality group than in the non-mortality group, as shown in Table 2. One study emphasized the effects of organ failure type and sequence on patient prognosis. Patients with AP with the cardiovascular or respiratory system as the first failing organ may have worse outcomes than those with the renal system involved.[7] Acute respiratory distress syndrome (ARDS), characterized by low blood oxygen levels, is the main cause of respiratory failure.[14] In the present study, according to the SHAP plot, the lower the Spo2_min, the easier it was for the model to predict a patient's death (Fig. 5). ARDS severity is closely associated with patient prognosis, explaining why the model makes such predictions. Given the serious consequences of ARDS, its prevention and management should be an important part of clinical practice

RDW, an important parameter of red blood cells, reflects the degree of variation in red blood cell volume.[15] Previous studies have shown that RDW is associated with many diseases, such as cardiovascular disease, venous thromboembolism, cancer, kidney disease, and liver disease.[15] In a large community-based sample, a higher RDW was associated with increased mortality risk.[16] Previous studies have explored the relationship between RDW and AP. Patients with RDW higher than the upper limit of normal were more likely to be admitted to the ICU, with a risk ratio of 3.5.[17] Zhang et al found that RDW has good predictive value for the prognosis of patients with AP, even better than APACHE-II and SOFA scores[18]. Consistent with the previous study, the present study showed that patients with a wider RDW had a higher mortality risk.

In the present study, BUN levels significantly differed between the survivor and nonsurvivor groups in the MIMIC-IV and eICU-CRD. The GNB model suggests that increased BUN levels indicate a higher mortality risk. Previous studies have found that BUN and changes in BUN levels are predictors of disease severity and death.[19–22] The AUC of BUN 24 hour after admission for predicting SAP was 0.80, higher than that at admission (AUC = 0.75). Furthermore, a BUN level of 8.3 mmol/L 24 hour after admission was observed as the optimal cutoff point for predicting SAP.[21] BUN is an excellent indicator reflecting the severity of AP, possibly because it reflects the intravascular volume depletion in the early stage of AP, the effect of liquid resuscitation, and the change in renal function due to AP.[21,23] The occurrence of AKI also reflects the impact of renal function on patient prognosis. As shown in the Fig. 5, patients without AKI have a lower risk of death.

Furthermore, several prognostic scoring systems, including Ranson, Glasgow, multiple organ system score, systemic inflammatory response syndrome, APACHE-II, and bedside index for severe acute pancreatitis, have been used to predict the severity and prognosis of AP patients.[8,24] SOFA is widely used to assess the condition of critically ill patients in the ICU.[25] Therefore, the SOFA score was used for prediction and was considered an important predictor in the GNB model. According to the SHAP plot, patients with a high SOFA score were considered to be at a high mortality risk (Fig. 5). In recent years, researchers have explored new methods of predicting the prognosis of AP. Based on the MIMIC-III and eICU-CRD, Xu et al established a nomogram and found that total bilirubin and albumin levels are important predictors of in-hospital mortality of patients with AP.[26] A previous study used a feed-forward ANN to predict the development of severe AP and fatal outcomes and found that the ANN showed greater accuracy than APACHE-II and Glasgow scores.[27] In another study, 5 logistic regression and 3 ANN models were constructed using data from 234 patients with SAP and were validated in 60 patients. The highest AUC of the logistic regression model was 0.862, whereas the ANN had an AUC of 0.847.[12] Ding et al created an ANN model based on the MIMIV-III database to predict in-hospital mortality. The ANN had an AUC of 0.769, better than that of logistic regression, Ranson score, and SOFA score. In their study, total bilirubin, amylase, ALT, and creatinine levels were the top 4 predictors of in-hospital mortality.[28] However, the sample sizes of these studies were relatively small, which may have impaired the performance of the machine-learning model. Hameed et al constructed 6 machine-learning methods based on MIMIC-III, MIMIC-IV, and eICU-CRD. Random Forest classifier demonstrated the best performance.[29] All patients with AP were enrolled in their study. Notably, most patients can recover quickly, with a low mortality risk of death.[6] Therefore, in the present study, we only enrolled patients with AP admitted to the ICU to construct a machine-learning model and perform external verification. The GNB model performed well on MIMIC-IV and eICU-CRD.

However, this study had some limitations. First, some important variables were eliminated due to many missing data points. Second, this was a retrospective study; therefore, the model should be evaluated in prospective studies.

TABLE 2. Baseline Characteristic of the Cohorts

	MIMIC-IV Cohort					eICU Cohort			
Variables	**ALL (n = 856)**	**Survivors 0 (n = 745)**	**Nonsurvivors 1 (n = 111)**	***P***		**ALL (n = 425)**	**Survivors 0 (n = 392)**	**Nonsurvivors 1 (n = 33)**	***P***
Demographics									
Age, median[IQR]	59.000 [46.000, 72.000]	57.000 [46.000, 71.000]	69.000 [58.000, 80.000]	< 0.001	—	53.000 [42.000, 63.000]	52.000 [42.000, 63.000]	62.000 [54. 000, 78.000]	< 0.001
Gender, n(%)									
Female	363 (42.407)	311 (41.745)	52 (46.847)	0.31	Female	171 (40.235)	157 (40.051)	14 (42.424)	0.789
Male	493 (57.593)	434 (58.255)	59 (53.153)	—	Male	254 (59.765)	235 (59.949)	19 (57.576)	—
Weight, median[IQR]	81.800 [70.000, 98.900]	81.800 [70.000, 99.000]	81.700 [68.700, 96.000]	0.559	—	84.000 [71.000, 99.800]	84.000 [71.600, 99.800]	84.200 [69.800, 99.300]	0.938
Height ,median[IQR]	169.745 [162.560, 177.010]	169.970 [163.000, 177.380]	168.000 [160.020, 175.845]	0.433	—	172.700 [165.000, 179.000]	172.700 [165.000, 179.000]	172.700 [165.000, 178.000]	0.959
Vital signs									
SpO2_mean ,median [IQR], %	96.292 [94.905, 97.780]	96.348 [94.960, 97.846]	95.692 [94.148, 97.250]	0.001	—	96.200 [95.000, 97.538]	96.192 [95.000, 97.500]	96.400 [95.143, 98.262]	0.634
SpO2_min ,median [IQR], %	92.000 [90.000, 94.000]	92.000 [90.000, 94.000]	90.000 [85.000, 92.000]	< 0.001	—	92.000 [89.000, 95.000]	92.000 [90.000, 95.000]	91.000 [87.000, 94.000]	0.064
TP_max,median [IQR], °C	37.500 [37.060, 38.170]	37.500 [37.110, 38.170]	37.220 [36.940, 37.940]	0.002	—	37.500 [37.100, 38.100]	37.500 [37.100, 38.100]	37.700 [37.100, 38.300]	0.425
TP_min, median [IQR], °C	36.500 [36.170, 36.780]	36.560 [36.220, 36.780]	36.330 [35.670, 36.560]	< 0.001	—	36.500 [36.200, 36.800]	36.500 [36.200, 36.800]	36.400 [36.200, 36.800]	0.378
MBP_mean, median [IQR], mm Hg	80.300 [72.185, 91.087]	81.440 [72.885, 91.667]	72.683 [67.919, 80.304]	< 0.001	—	85.208 [74.033, 97.063]	86.000 [75.179, 98.693]	71.480 [67.217, 81.733]	< 0.001
DBP_mean, median [IQR], mm Hg	66.600 [58.222, 76.410]	67.351 [59.040, 77.625]	59.560 [53.250, 66.600]	< 0.001	—	71.462 [63.087, 82.118]	72.240 [64.000, 82.904]	59.839 [56.714, 68.241]	< 0.001
SBP_mean, median [IQR], mm Hg	119.040 [107.758, 134.467]	120.560 [108.990, 135.609]	109.800 [101.479, 117.778]	< 0.001	—	125.466 [110.102, 141.094]	127.219 [112.500, 141.500]	106.714 [98.222, 116.579]	< 0.001
RR_mean ,median [IQR]	20.560 [17.778, 23.913]	20.259 [17.462, 23.560]	22.220 [19.963, 25.125]	< 0.001	—	20.292 [17.788, 23.902]	20.197 [17.639, 23.725]	21.114 [19.500, 26.425]	0.076
RR_max, median [IQR]	29.000 [25.000,34.000]	29.000 [25.000, 34.000]	32.000 [27.000, 36.000]	< 0.001	—	29.000 [24.000, 35.000]	28.500 [24.000, 34.000]	33.000 [29.000, 38.000]	0.002
RR_min, median [IQR]	14.000 [11.000, 16.000]	13.500 [11.000, 16.000]	14.000 [11.000, 16.000]	0.812	—	14.000 [12.000, 18.000]	14.333 [12.000, 18.000]	14.000 [10.000, 18.000]	0.378
HR_mean, median [IQR]	95.069 [82.154, 108.229]	95.125 [81.560, 108.038]	93.038 [85.407, 108.567]	0.732	—	99.087 ± 17.893	98.792 ± 17.706	102.591 ± 19.646	0.242
HR_max, median [IQR]	115.000 [99.000, 130.000]	115.000 [99.000, 128.000]	118.000 [103.000, 132.000]	0.235	—	117.429 [102.000, 133.000]	117.000 [102.000, 133.000]	125.000 [107.000, 152.000]	0.099
HR_min, median [IQR]	79.000 [68.000, 94.000]	79.000 [68.000, 94.000]	78.000 [68.000, 93.000]	0.83	—	84.632 ± 18.529	84.540 ± 18.396	85.727 ± 20.000	0.724
Laboratory tests									
Sodium_max, median[IQR], mmol/L	140.000 [137.000, 143.000]	140.000 [137.000, 143.000]	140.000 [136.000, 143.000]	0.657	—	139.000 [137.000, 143.000]	139.000 [137.000, 143.000]	139.000 [137.000, 142.000]	0.938
Sodium_min, median [IQR], mmol/L	136.000 [133.000, 140.000]	136.000 [133.000, 140.000]	137.000 [132.000, 140.000]	0.876	—	135.000 [131.000, 138.000]	135.000 [131.000, 138.000]	136.000 [132.000, 137.000]	0.823
Chloride_max, median [IQR], mmol/L	107.000 [102.000, 111.000]	106.000 [103.000, 111.000]	109.000 [102.000, 112.000]	0.261	—	107.000 [102.000, 111.000]	107.000 [102.000, 111.000]	106.000 [102.000, 110.000]	0.271
Chloride_min, median [IQR], mmol/L	102.000 [97.000, 106.000]	102.000 [97.000, 106.000]	102.000 [97.000, 107.000]	0.985	—	99.000 [95.000, 104.000]	99.000 [94.000, 104.000]	100.000 [96.000, 103.000]	0.702
Bicarbonate_max, median [IQR], mmol/L	23.000 [20.000, 26.000]	24.000 [21.000, 26.000]	22.000 [19.000, 25.000]	< 0.001	—	24.400 [22.000, 27.000]	24.667 [21.000, 27.000]	23.000 [22.000, 27.000]	0.576
Bicarbonate_min, median [IQR], mmol/L	20.000 [16.000, 23.000]	20.000 [17.000, 23.000]	17.000 [13.000, 21.000]	< 0.001	—	19.000 [15.333, 23.000]	20.000 [16.000, 23.000]	19.000 [14.000, 22.000]	0.389
Potassium_max, median [IQR], mmol/L	4.400 [4.000, 5.000]	4.300 [4.000, 4.900]	4.600 [4.200, 5.400]	< 0.001	—	4.400 [4.000, 4.900]	4.300 [3.950, 4.900]	4.700 [4.100, 5.400]	0.052
Potassium_min, median [IQR], mmol/L	3.700 [3.400, 4.100]	3.700 [3.400, 4.100]	3.900 [3.500, 4.300]	0.002	—	3.600 [3.200, 3.900]	3.600 [3.200, 3.900]	3.600 [3.400, 4.100]	0.103
Calcium_max, median [IQR], mg/dL	8.300 [7.800, 8.800]	8.300 [7.800, 8.800]	8.300 [7.800, 8.900]	0.61	—	8.800 [8.000, 9.400]	8.800 [8.100, 9.400]	8.400 [7.800, 9.400]	0.222
Calcium_min, median [IQR], mg/dL	7.700 [7.100, 8.200]	7.700 [7.100, 8.200]	7.500 [6.800, 8.000]	0.016	—	7.500 [6.900, 8.200]	7.600 [6.900, 8.200]	7.400 [7.000, 7.900]	0.607
Aniongap_max, median [IQR], mmol/L	17.000 [14.000, 21.000]	17.000 [14.000, 20.000]	19.000 [15.000, 24.000]	< 0.001	—	16.000 [12.000, 21.571]	16.000 [12.000, 21.000]	18.000 [13.100, 23.000]	0.369
Aniongap_min, median [IQR], mmol/L	13.000 [11.000, 16.000]	13.000 [11.000, 15.000]	16.000 [12.000, 17.000]	< 0.001	—	10.000 [7.000, 13.000]	10.000 [7.000, 13.000]	11.000 [8.000, 15.800]	0.087
Creatinine_max, median [IQR], mg/dL	1.200 [0.800, 2.200]	1.100 [0.800, 2.000]	2.000 [1.100, 3.700]	< 0.001	—	1.290 [0.820, 2.300]	1.250 [0.810, 2.080]	2.470 [1.280, 4.630]	< 0.001
BUN_max, median [IQR], mg/dL	23.000 [14.000, 40.000]	20.000 [13.000, 36.000]	42.000 [26.000, 64.000]	< 0.001	—	20.000 [12.000, 35.000]	19.000 [12.000, 34.000]	35.000 [22.000, 60.000]	< 0.001

Albumin_min, median [IQR], g/dL	3.000 [2.500,3.400]	3.000 [2.600,3.500]	2.800 [2.200,3.200]	< 0.001	—	2.798 ± 0.699	2.839 ± 0.695	2.314 ± 0.548	< 0.001
Bilirubin_total_max, median [IQR], mg/dL	1.300 [0.600, 3.200]	1.200 [0.600, 3.000]	2.200 [0.900, 5.400]	< 0.001	—	1.300 [0.700, 2.600]	1.210 [0.700, 2.400]	3.000 [0.950, 7.400]	0.009
Ast_max, median [IQR], Units/L	86.000 [39.000, 220.000]	80.000 [37.000, 200.000]	129.000 [53.000, 300.000]	< 0.001	—	84.000 [36.000, 213.000]	83.000 [36.000, 211.000]	135.000 [35.000, 290.000]	0.293
ALT_max, median [IQR], Units/L	60.000 [26.000, 167.000]	60.000 [26.000, 163.000]	59.000 [30.000, 204.000]	0.375	—	56.000 [29.000, 132.000]	56.000 [30.000, 130.000]	51.000 [22.000, 138.000]	0.716
Platelets_min, median [IQR], 10^9/L	167.000 [111.000,234.000]	170.000 [116.000, 237.000]	137.000 [76.000, 216.000]	0.004	—	162.000 [112.000, 225.000]	161.000 [112.000, 224.000]	165.000 [115.000, 233.000]	0.932
WBC_max, median [IQR], 10^9/L	14.100 [10.000,19.800]	14.000 [9.900, 19.500]	16.000 [11.700, 22.200]	0.006	—	14.800 [10.300, 19.950]	14.540 [10.100, 19.500]	17.800 [14.000, 24.350]	0.009
WBC_min, median [IQR], 10^9/L	10.700 [7.200,15.700]	10.500 [7.100, 15.500]	12.400 [7.700, 16.900]	0.058	—	9.900 [6.600, 13.600]	9.800 [6.600, 13.300]	11.600 [6.800, 16.500]	0.151
RDW_max, median [IQR], 10^9/L	14.900 [13.900,16.400]	14.800 [13.800, 16.000]	16.200 [14.800, 18.400]	< 0.001	—	15.000 [13.900, 16.200]	14.900 [13.800, 16.000]	16.400 [15.000, 19.000]	< 0.001
RBC_min, median [IQR], 10^9/L	3.390 [2.910,3.920]	3.440 [2.970,3.940]	3.050 [2.540,3.680]	< 0.001	—	3.812 ± 0.813	3.857 ± 0.799	3.272 ± 0.783	< 0.001
Hemoglobin_min, mean (± SD), g/dL	10.413 ± 2.279	10.558 ± 2.253	9.441 ± 2.214	< 0.001	—	11.800 [10.000, 13.500]	12.000 [10.000, 13.600]	10.000 [8.200, 12.000]	< 0.001
Hematocrit_min, mean (± SD), %	31.304 ± 6.638	31.658 ± 6.570	28.923 ± 6.600	< 0.001	—	35.500 [29.900, 39.700]	35.800 [30.200, 40.000]	30.500 [25.400, 36.300]	0.002
Comorbidities									
Hypertension, n(%)									
No	314 (36.682)	279 (37.450)	35 (31.532)	0.227	No	217 (51.059)	206 (52.551)	11 (33.333)	0.034
Yes	542 (63.318)	466 (62.550)	76 (68.468)	—	Yes	208 (48.941)	186 (47.449)	22 (66.667)	—
Diabetes, n(%)									
No	586 (68.458)	511 (68.591)	75 (67.568)	0.829	No	279 (65.647)	257 (65.561)	22 (66.667)	0.898
Yes	270 (31.542)	234 (31.409)	36 (32.432)	—	Yes	146 (34.353)	135 (34.439)	11 (33.333)	—
Complication									
AKI									
No	316 (36.916)	299 (40.134)	17 (15.315)	< 0.001	No	253 (59.529)	245 (62.500)	8 (24.242))	< 0.001
Yes	540 (63.084)	446 (59.866)	94 (84.685)	—	Yes	172 (40.471)	147 (37.500)	25 (75.758)	—
Prognostic scoring system									
OASIS, median[IQR]	32.000 [26.000, 39.000]	31.000 [26.000, 37.000]	41.000 [34.000, 47.000]	< 0.001	—	22.000 [17.000, 30.000]	22.000 [16.000, 29.000]	33.000 [30.000, 39.000]	< 0.001
SOFA, median[IQR]	5.000 [3.000, 9.000]	5.000 [2.000, 8.000]	9.000 [6.000, 13.000]	< 0.001	—	5.000 [3.000, 8.000]	5.000 [3.000, 7.000]	10.000 [8.000, 12.000]	< 0.001
Treatment									
Mechanical ventilation, n(%)									
NO	521 (60.864)	470 (63.087)	51 (45.946)	< 0.001	No	306 (72.000)	295 (75.255)	11 (33.333)	< 0.001
Yes	335 (39.136)	275 (36.913)	60 (54.054)	—	Yes	119 (28.000)	97 (24.745)	22 (66.667)	—
RRT, n(%)									
NO	825 (96.379)	728 (97.718)	97 (87.387)	< 0.001	No	380 (89.412)	358 (91.327)	22 (66.667)	< 0.001
Yes	31 (3.621)	17 (2.282)	14 (12.613)	—	Yes	45 (10.588)	34 (8.673)	11 (33.333)	—

AKI indicates acute kidney injury; ALT, aminotransferase alanine; AST, aminotransferase aspartate; BUN, blood urea nitrogen; DBP, diastolic blood pressure; HR, heart rate; max, maximum; MBP, mean blood pressure; Min, minimum; OASIS, Oxford acute severity of illness score; RBC, red blood cells; RDW, red cell distribution width; RR, respiratory rate; RRT, renal replacement therapy, SBP, systolic blood pressure; SOFA, Sequential Organ Failure Assessment; SpO2, saturation of blood oxygen; TP, temperature; WBC, white blood cells.

CONCLUSION

In conclusion, the GNB model demonstrated excellent performance and generalizability in predicting mortality in ICU patients with AP. Therefore, it can identify patients at a high mortality risk.

REFERENCES

1. Sendler M, Algül H. Pathogenesis of acute pancreatitis. *Internist (Berl)*. 2021;62:1034–1043.
2. Lee PJ, Papachristou GI. New insights into acute pancreatitis. *Nat Rev Gastroenterol Hepatol*. 2019;16:479–496.
3. Mederos MA, Reber HA, Girgis MD. Acute Pancreatitis: A Review. *JAMA*. 2021;325:382–390.
4. Forsmark CE, Vege SS, Wilcox CM. Acute Pancreatitis. *N Engl J Med*. 2016;375:1972–1981; united States: Massachusetts Medical Society.
5. Banks PA, Bollen TL, Dervenis C, et al. Classification of acute pancreatitis–2012: revision of the Atlanta classification and definitions by international consensus. *Gut*. 2013;62:102–111.
6. Boxhoorn L, Voermans RP, Bouwense SA, et al. Acute pancreatitis. *Lancet*. 2020;396:726–734.
7. Machicado JD, Gougol A, Tan X, et al. Mortality in acute pancreatitis with persistent organ failure is determined by the number, type, and sequence of organ systems affected. *United European Gastroenterol J*. 2021;9:139–149.
8. Khanna AK, Meher S, Prakash S, et al. Comparison of Ranson, Glasgow, MOSS, SIRS, BISAP, APACHE-II, CTSI Scores, IL-6, CRP, and Procalcitonin in Predicting Severity, Organ Failure, Pancreatic Necrosis, and Mortality in Acute Pancreatitis. *HPB Surg*. 2013;2013:367581.
9. Tee YS, Fang HY, Kuo IM, et al. Serial evaluation of the SOFA score is reliable for predicting mortality in acute severe pancreatitis. *Medicine (Baltimore)*. 2018;97:e9654.
10. Tan Q, Hu C, Chen Z, et al. Growth differentiation factor 15 is an early predictor for persistent organ failure and mortality in acute pancreatitis. *Pancreatology*. 2022;22:200–209.
11. Li J, Chen Z, Li L, et al. Interleukin-6 is better than C-reactive protein for the prediction of infected pancreatic necrosis and mortality in patients with acute pancreatitis. *Front Cell Infect Microbiol*. 2022;12:933221.
12. Halonen KI, Leppäniemi AK, Lundin JE, et al. Predicting fatal outcome in the early phase of severe acute pancreatitis by using novel prognostic models. *Pancreatology*. 2003;3:309–315.
13. Schepers NJ, Bakker OJ, Besselink MG, et al. Impact of characteristics of organ failure and infected necrosis on mortality in necrotising pancreatitis. *Gut*. 2019;68:1044–1051.
14. Ranieri VM, Rubenfeld GD, Thompson BT, et al. Acute respiratory distress syndrome: the Berlin Definition. *JAMA*. 2012;307:2526–2533.
15. Salvagno GL, Sanchis-Gomar F, Picanza A, et al. Red blood cell distribution width: A simple parameter with multiple clinical applications. *Crit Rev Clin Lab Sci*. 2015;52:86–105.
16. Perlstein TS, Weuve J, Pfeffer MA, et al. Red blood cell distribution width and mortality risk in a community-based prospective cohort. *Arch Intern Med*. 2009;169:588–594.
17. O'Connell RM, Boland MR, O'Driscoll J, et al. Red cell distribution width and neutrophil to lymphocyte ratio as predictors of outcomes in acute pancreatitis: a retrospective cohort study. *Int J Surg*. 2018;55:124–127.
18. Zhang FX, Li ZL, Zhang ZD, et al. Prognostic value of red blood cell distribution width for severe acute pancreatitis. *World J Gastroenterol*. 2019;25:4739–4748.
19. Vitale DS, Hornung L, Lin TK, et al. Blood Urea Nitrogen Elevation Is a Marker for Pediatric Severe Acute Pancreatitis. *Pancreas*. 2019;48:363–366.
20. Pando E, Alberti P, Mata R, et al. Early Changes in Blood Urea Nitrogen (BUN) can predict mortality in acute pancreatitis: Comparative Study between BISAP Score, APACHE-II, and Other Laboratory Markers-A Prospective Observational Study. *Can J Gastroenterol Hepatol*. 2021; 2021:6643595.
21. Lin S, Hong W, Basharat Z, et al. Blood urea nitrogen as a predictor of severe acute pancreatitis based on the revised atlanta criteria: timing of measurement and cutoff points. *Can J Gastroenterol Hepatol*. 2017;2017:9592831.
22. Wu BU, Johannes RS, Sun X, et al. Early changes in blood urea nitrogen predict mortality in acute pancreatitis. *Gastroenterology*. 2009;137:129–135.
23. Wu BU, Johannes RS, Sun X, et al. The early prediction of mortality in acute pancreatitis: a large population-based study. *Gut*. 2008;57:1698–1703.
24. Halonen KI, Pettilä V, Leppäniemi AK, et al. Multiple organ dysfunction associated with severe acute pancreatitis. *Crit Care Med*. 2002;30:1274–1279.
25. Lambden S, Laterre PF, Levy MM, et al. The SOFA score-development, utility and challenges of accurate assessment in clinical trials. *Crit Care*. 2019;23:374.
26. Xu X, Ai F, Huang M. Deceased serum bilirubin and albumin levels in the assessment of severity and mortality in patients with acute pancreatitis. *Int J Med Sci*. 2020;17:2685–2695.
27. Mofidi R, Duff MD, Madhavan KK, et al. Identification of severe acute pancreatitis using an artificial neural network. *Surgery*. 2007;141:59–66.
28. Ding N, Guo C, Li C, et al. An Artificial Neural Networks Model for Early Predicting In-Hospital Mortality in Acute Pancreatitis in MIMIC-III. *Biomed Res Int*. 2021;2021: 6638919.
29. Hameed M, Alamgir Z. Improving mortality prediction in Acute Pancreatitis by machine learning and data augmentation. *Comput Biol Med*. 2022;150:106077.